Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:

- highly multivariate, with many important predictors and response variables,
- geographically referenced, and often presented as maps, and
- temporally correlated, as in longitudinal or other time series structures.

⇒ motivates hierarchical modeling and data analysis for complex spatial (and spatiotemporal) data sets.
- point-referenced data, where \( Y(s) \) is a random vector at a location \( s \in \mathbb{R}^r \), where \( s \) varies continuously over \( D \), a fixed subset of \( \mathbb{R}^r \) that contains an \( r \)-dimensional rectangle of positive volume;

- areal data, where \( D \) is again a fixed subset (of regular or irregular shape), but now partitioned into a finite number of areal units with well-defined boundaries;

- point pattern data, where now \( D \) is itself random; its index set gives the locations of random events that are the spatial point pattern. \( Y(s) \) itself can simply equal 1 for all \( s \in D \) (indicating occurrence of the event), or possibly give some additional covariate information (producing a marked point pattern process).

First step in analyzing data

First Law of Geography: Mean + Error

Mean: first-order behavior
Introduction to spatial data and models

Exploration of spatial data

First step in analyzing data

First Law of Geography: Mean + Error

Mean: first-order behavior

Error: second-order behavior (covariance function)

EDA tools examine both first and second order behavior

Preliminary displays: Simple locations to surface displays

Scallops Sites

Spatial surface observed at finite set of locations

\[ S = \{ s_1, s_2, \ldots, s_n \} \]

Tessellate the spatial domain (usually with data locations as vertices)

Fit an interpolating polynomial:

\[ f(s) = \sum_i w_i(S; s) f(s_i) \]

“Interpolate” by reading off \( f(s_0) \).

Issues:

- Sensitivity to tessellations
- Choices of multivariate interpolators
- Numerical error analysis
Point-level modelling refers to modelling of spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing).

**Fundamental concept:** Data from a spatial process \{Y(s) : s ∈ D\}, where \(D\) is a fixed subset in Euclidean space.

**Example:** \(Y(s)\) is a pollutant level at site \(s\).

**Conceptually:** Pollutant level exists at all possible sites.

**Practically:** Pollutant level exists at all possible sites.
Suppose our spatial process has a mean, $\mu(s) = E(Y(s))$, and that the variance of $Y(s)$ exists for all $s \in D$.

- **Strong stationarity**: If for any given set of sites, and any displacement $h$, the distribution of $(Y(s_1), \ldots, Y(s_n))$ is the same as $(Y(s_1 + h), \ldots, Y(s_n + h))$.
- **Weak stationarity**: Constant mean $\mu(s) = \mu$, and $\text{Cov}(Y(s), Y(s+h)) = C(h)$: the covariance depends only upon the displacement (or separation) vector.

Suppose our spatial process has a mean, $\mu(s) = E(Y(s))$, and that the variance of $Y(s)$ exists for all $s \in D$.

- **Strong stationarity**: If for any given set of sites, and any displacement $h$, the distribution of $(Y(s_1), \ldots, Y(s_n))$ is the same as $(Y(s_1 + h), \ldots, Y(s_n + h))$.
- **Weak stationarity**: Constant mean $\mu(s) = \mu$, and $\text{Cov}(Y(s), Y(s+h)) = C(h)$: the covariance depends only upon the displacement (or separation) vector.

Suppose our spatial process has a mean, $\mu(s) = E(Y(s))$, and that the variance of $Y(s)$ exists for all $s \in D$.

- **Strong stationarity**: If for any given set of sites, and any displacement $h$, the distribution of $(Y(s_1), \ldots, Y(s_n))$ is the same as $(Y(s_1 + h), \ldots, Y(s_n + h))$.
- **Weak stationarity**: Constant mean $\mu(s) = \mu$, and $\text{Cov}(Y(s), Y(s+h)) = C(h)$: the covariance depends only upon the displacement (or separation) vector.

- **Strong stationarity implies weak stationarity**
- **The process is** Gaussian if $Y = (Y(s_1), \ldots, Y(s_n))$ has a multivariate normal distribution.

- **Strong stationarity implies weak stationarity**
- The process is Gaussian if $Y = (Y(s_1), \ldots, Y(s_n))$ has a multivariate normal distribution.
- For Gaussian processes, strong and weak stationarity are equivalent.

- **Point-level modelling** refers to modelling of spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing).
- **Fundamental concept**: Data from a spatial process $\{Y(s) : s \in D\}$, where $D$ is a fixed subset in Euclidean space.
- **Example**: $Y(s)$ is a pollutant level at site $s$
- **Conceptually**: Pollutant level exists at all possible sites
- **Practically**: Data will be a partial realization of a spatial process – observed at $\{s_1, \ldots, s_n\}$
- **Statistical objectives**: Inference about the process $Y(s)$; predict at new locations.

**Strong stationarity implies weak stationarity**

The process is Gaussian if $Y = (Y(s_1), \ldots, Y(s_n))$ has a multivariate normal distribution.

- **Strong stationarity**: If for any given set of sites, and any displacement $h$, the distribution of $(Y(s_1), \ldots, Y(s_n))$ is the same as $(Y(s_1 + h), \ldots, Y(s_n + h))$.
- **Weak stationarity**: Constant mean $\mu(s) = \mu$, and $\text{Cov}(Y(s), Y(s+h)) = C(h)$: the covariance depends only upon the displacement (or separation) vector.

- **Strong stationarity implies weak stationarity**
- **The process is** Gaussian if $Y = (Y(s_1), \ldots, Y(s_n))$ has a multivariate normal distribution.
- For Gaussian processes, strong and weak stationarity are equivalent.
Variograms

- Suppose we assume $E[Y(\mathbf{s} + h) - Y(\mathbf{s})] = 0$ and define
  
  $$
  E[Y(\mathbf{s} + h) - Y(\mathbf{s})]^2 = Var(Y(\mathbf{s} + h) - Y(\mathbf{s})) = 2\gamma(h).
  $$

  This is sensible if the left hand side depends only upon $h$. Then we say the process is intrinsically stationary.

- $\gamma(h)$ is called the semivariogram and $2\gamma(h)$ is called the variogram.

Note that intrinsic stationarity defines only the first and second moments of the differences $Y(\mathbf{s} + h) - Y(\mathbf{s})$. It says nothing about the joint distribution of a collection of variables $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$, and thus provides no likelihood.

Intrinsic Stationarity and Ergodicity

- Relationship between $\gamma(h)$ and $C(h)$:
  
  $$
  2\gamma(h) = Var(Y(\mathbf{s} + h)) + Var(Y(\mathbf{s})) - 2Cov(Y(\mathbf{s} + h), Y(\mathbf{s}))
  = C(\mathbf{0}) + C(\mathbf{0}) - 2C(h)
  = 2[C(\mathbf{0}) - C(h)].
  $$

- Easy to recover $\gamma$ from $C$. The converse needs the additional assumption of ergodicity: $\lim_{|\mathbf{u}| \to \infty} C(\mathbf{u}) = 0$. 

- $\gamma(\mathbf{h})$ is called the semivariogram and $2\gamma(\mathbf{h})$ is called the variogram.

Intrinsic Stationarity and Ergodicity

- Relationship between $\gamma(h)$ and $C(h)$:
  
  $$
  2\gamma(h) = Var(Y(\mathbf{s} + h)) + Var(Y(\mathbf{s})) - 2Cov(Y(\mathbf{s} + h), Y(\mathbf{s}))
  = C(\mathbf{0}) + C(\mathbf{0}) - 2C(h)
  = 2[C(\mathbf{0}) - C(h)].
  $$

- Easy to recover $\gamma$ from $C$. The converse needs the additional assumption of ergodicity: $\lim_{|\mathbf{u}| \to \infty} C(\mathbf{u}) = 0$.

- So $\lim_{|\mathbf{u}| \to \infty} \gamma(\mathbf{u}) = C(\mathbf{0})$, and we can recover $C$ from $\gamma$ as long as this limit exists.

  $$
  C(h) = \lim_{|\mathbf{u}| \to \infty} \gamma(\mathbf{u}) - \gamma(h).
  $$

Intrinsic Stationarity and Ergodicity
- When $\gamma(h)$ or $C(h)$ depends upon the separation vector only through the distance $|h|$, we say that the process is isotropic. In that case, we write $\gamma(|h|)$ or $C(|h|)$. Otherwise we say that the process is anisotropic.

- If the process is intrinsically stationary and isotropic, it is also called homogeneous.

Some common isotropic variograms

<table>
<thead>
<tr>
<th>Model</th>
<th>Variogram, $\gamma(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t &amp; \text{if } t &gt; 0 \ 0 &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>Spherical</td>
<td>$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[ \frac{3}{2}</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi t)) &amp; \text{if } t &gt; 0 \ 0 &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>Powered</td>
<td>$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-</td>
</tr>
<tr>
<td>Matérn</td>
<td>$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[ 1 - (1 + \phi t) e^{-\phi t} \right] &amp; \text{if } t &gt; 0 \ 0 &amp; \text{o/w} \end{cases}$</td>
</tr>
</tbody>
</table>

Examples: Spherical Variogram

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[ \frac{3}{2} |\phi t| - \frac{1}{2} (\phi t)^3 \right] & \text{if } 0 < t \leq 1/\phi \\ \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \end{cases}$$

- While $\gamma(0) = 0$ by definition, $\gamma(0^+) \equiv \lim_{t \to 0^+} \gamma(t) = \tau^2$; this quantity is the nugget.
Examples: Spherical Variogram

\[ \gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \\ \frac{1}{2} \sigma^2 t^2 - \frac{1}{4} (\phi t)^2 & \text{if } 0 < t \leq 1/\phi \\ 0 & \text{if } t = 0. \end{cases} \]

- While \( \gamma(0) = 0 \) by definition, \( \gamma(0^+) \equiv \lim_{t \to 0^+} \gamma(t) = \tau^2; \) this quantity is the nugget.
- \( \lim_{t \to \infty} \gamma(t) = \tau^2 + \sigma^2; \) this asymptotic value of the semivariogram is called the sill. (The sill minus the nugget, \( \sigma^2 \) in this case, is called the partial sill.)

Some common isotropic covariograms

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariance function, ( C(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>( C(t) ) does not exist</td>
</tr>
<tr>
<td>Spherical</td>
<td>( C(t) = \begin{cases} \tau^2 &amp; \text{if } t \geq 1/\phi \ \sigma^2 t^2 &amp; \text{if } 0 &lt; t \leq 1/\phi \ \sigma^2 &amp; \text{otherwise} \end{cases} )</td>
</tr>
<tr>
<td>Exponential</td>
<td>( C(t) = \begin{cases} \sigma^2 \exp(-\phi t) &amp; \text{if } t &gt; 0 \ \tau^2 + \sigma^2 &amp; \text{otherwise} \end{cases} )</td>
</tr>
<tr>
<td>Powered exponential</td>
<td>( C(t) = \begin{cases} \sigma^2 \exp(-\phi t^p) &amp; \text{if } t &gt; 0 \ \tau^2 + \sigma^2 &amp; \text{otherwise} \end{cases} )</td>
</tr>
<tr>
<td>Matérn at ( \nu = 3/2 )</td>
<td>( C(t) = \begin{cases} \sigma^2 (1 + \phi t) \exp(-\phi t) &amp; \text{if } t &gt; 0 \ \tau^2 + \sigma^2 &amp; \text{otherwise} \end{cases} )</td>
</tr>
</tbody>
</table>

Notes on exponential model

\[ C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0. \end{cases} \]

- We define the effective range, \( t_0 \), as the distance at which this correlation has dropped to only 0.05. Setting \( \exp(-\phi t_0) \) equal to this value we obtain \( t_0 \approx 3/\phi \), since \( \log(0.05) \approx -3 \).
Introduction to spatial data and models

Notes on exponential model

\[ C(t) = \begin{cases} \sigma^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0 \end{cases} \]

- We define the effective range, \( t_0 \), as the distance at which this correlation has dropped to only 0.05. Setting \( \exp(-\phi t_0) = 0.05 \), we obtain \( t_0 \approx 3/\phi \), since \( \log(0.05) \approx -3 \).
- Finally, the form of \( C(t) \) shows why the nugget \( \sigma^2 \) is often viewed as a “nonspatial effect variance,” and the partial sill \( (\sigma^2) \) is viewed as a “spatial effect variance.”

The Matérn Correlation Function

- Much of statistical modelling is carried out through correlation functions rather than variograms

\[ C(t) = \begin{cases} \frac{\sigma^2}{\sqrt{2\nu\pi\tau}} \left( \frac{s}{\nu\tau} \right)^\nu K_\nu \left( \frac{s}{\nu\tau} \right) & \text{if } t > 0 \\ \sigma^2 & \text{if } t = 0 \end{cases} \]

\( K_\nu \) is the modified Bessel function of order \( \nu \) (computationally tractable)

How do we select a variogram? Can the data really distinguish between variograms?

Empirical Variogram:

\[ \gamma(t) = \frac{1}{2N(t)} \sum_{s_i, s_j \in N(t)} (Y(s_i) - Y(s_j))^2 \]

where \( N(t) \) is the number of points such that \( ||s_i - s_j|| = t \) and \( |N(t)| \) is the number of points in \( N(t) \).
How do we select a variogram? Can the data really distinguish between variograms?

**Empirical Variogram:**

\[ \gamma(t) = \frac{1}{2|N(t)|} \sum_{s_i, s_j \in N(t)} (Y(s_i) - Y(s_j))^2 \]

where \( N(t) \) is the number of points such that \( |s_i - s_j| = t \) and \( |N(t)| \) is the number of points in \( N(t) \).

* Grid up the \( t \) space into intervals \( I_1 = (0, t_1), I_2 = (t_1, t_2), \) and so forth, up to \( I_K = (t_{K-1}, t_K) \). Representing \( t \) values in each interval by its midpoint, we define:

\[ N(t_k) = \{(s_i, s_j) : |s_i - s_j| \in I_k\}, k = 1, \ldots, K. \]